# Stappenplan 28 Aug

## Stap 1: extraheer factoren uit dataset

Je begint met de dataset die al is geladen en gefilterd (bijvoorbeeld met de Christiano-Fitzgerald filter).

Vervolgens gebruik je PCA (of PLS) om een matrix van 9 factoren te extraheren voor alle datapunten in je trainingsset (tot december 2019).

Deze factoren zullen de onderliggende structuur van je 66 variabelen representeren en dienen als input voor verdere voorspellingen.

### CHECK

*Dit is de terminal uitput na check stap 1:*

*>>> from data\_loader import load\_data, filter\_data*

*>>> from utils import standardize*

*>>> from factor\_model import DynamicFactorModel*

*>>>*

*>>> # Laad en filter data*

*>>> file\_path = 'C:/Thesis/03. Data/Final version data/Static.xlsx'*

*>>> df\_data = load\_data(file\_path)*

*>>> filtered\_df = filter\_data(df\_data)*

*>>>*

*>>> # Standaardiseer data*

*>>> Y\_train\_std = standardize(filtered\_df.values.T).T*

*>>>*

*>>> # Initialiseer en pas PCA toe*

*>>> model = DynamicFactorModel(filtered\_df, num\_factors=9)*

*>>> model.std\_data = Y\_train\_std.T*

*>>> model.apply\_pca()*

*>>>*

*>>> # Inspecteer de factors matrix*

*>>> print("Shape of factors matrix:", model.factors.shape)*

*Shape of factors matrix: (9, 347)*

*>>> print("First few rows of the factors matrix:\n", model.factors[:5, :])*

*First few rows of the factors matrix:*

*[[ 6.97008786 6.03887386 3.8322957 ... 2.61886857 5.43186564*

*3.59887794]*

*[ 5.99404811 5.2211467 5.24405989 ... 10.00381723 9.42566161*

*10.75307203]*

*[-1.12879002 0.77675893 -2.21537595 ... -1.96604757 -2.17006265*

*3.09251458]*

*[-2.0293299 -4.37194985 -4.57658607 ... -3.16489002 -3.56874705*

*-3.23346493]*

*[-0.4204558 -1.86658493 -1.00839733 ... 1.2301937 0.96019515*

*0.53709331]]*

*>>>*

## Stap 2: Schat model met ElasticNet

Gebruik de verkregen factoren uit stap 1 als onafhankelijke variabelen om een ElasticNet-regressiemodel te trainen. Dit model wordt getraind om de oorspronkelijke 66 variabelen te voorspellen op basis van de factoren.

### CHECK

*Dit is de terminal output na check stap 2:*

*>>> from data\_loader import load\_data, filter\_data*

*>>> from utils import standardize*

*>>> from factor\_model import DynamicFactorModel*

*>>>*

*>>> # Laad en filter data*

*>>> file\_path = 'C:/Thesis/03. Data/Final version data/Static.xlsx'*

*>>> df\_data = load\_data(file\_path)*

*>>> filtered\_df = filter\_data(df\_data)*

*>>>*

*>>> # Standaardiseer data*

*>>> Y\_train\_std = standardize(filtered\_df.values.T).T*

*>>>*

*>>> # Initialiseer het model en pas PCA toe*

*>>> model = DynamicFactorModel(filtered\_df, num\_factors=9)*

*>>> model.std\_data = Y\_train\_std.T*

*>>> model.apply\_pca()*

*>>>*

*>>> # Bereid de data voor de ElasticNet-training*

*>>> # Gebruik een subset van de data voor training om te testen*

*>>> train\_split\_index = int(model.factors.shape[1] \* 0.8)*

*>>> data\_train = Y\_train\_std[:, :train\_split\_index].T*

*>>> fac\_train = model.factors[:, :train\_split\_index].T*

*>>>*

*>>> # Train het ElasticNet-model*

*>>> B\_matrix, r2\_insample, intercept = model.enet\_fit(data\_train, fac\_train)*

*>>>*

*>>> # Inspecteer de resultaten*

*>>> print("Shape of B\_matrix (ElasticNet coefficients):", B\_matrix.shape)*

*Shape of B\_matrix (ElasticNet coefficients): (66, 9)*

*>>> print("First few coefficients of B\_matrix:\n", B\_matrix[:5, :5])*

*First few coefficients of B\_matrix:*

*[[ 0.04088399 0.08592133 -0.01551943 -0.02031759 0.00576677]*

*[ 0.01990409 0.04395976 -0.01622239 -0.00101789 -0.07826501]*

*[ 0.03348446 0.06306993 -0.00249369 -0.00724524 -0.03015038]*

*[ 0.0243287 0.07147932 -0.01749151 -0.01492411 -0.02055175]*

*[ 0.01565931 0.03460863 -0.01010059 0.02615383 -0.05804405]]*

*>>> print("In-sample R^2 value:", r2\_insample)*

*In-sample R^2 value: 0.7645413976926512*

*>>>*

## Stap 3: Voorspel factoren met Yule-Walker

Met het geschatte model en de factoren tot de laatste tijdstempel in de trainingsset, gebruik je de Yule-Walker vergelijking om de factoren voor de volgende tijdstempel t+1 te voorspellen.

Dit betekent dat je gebruik maakt van de geschatte autoregressieve parameters (phi) om toekomstige waarden van de factoren te schatten.

### CHECK

*Dit is de terminal output na check stap 3:*

*>>> # Begin met het importeren van de benodigde libraries en modules*

*>>> import pandas as pd*

*>>> from data\_loader import load\_data, filter\_data*

*>>> from utils import standardize*

*>>> from factor\_model import DynamicFactorModel*

*>>>*

*>>> # Laad en filter data*

*>>> file\_path = 'C:/Thesis/03. Data/Final version data/Static.xlsx'*

*>>> df\_data = load\_data(file\_path)*

*>>> filtered\_df = filter\_data(df\_data)*

*>>>*

*>>> # Definieer de trainingsperiode*

*>>> DATE\_TRAIN\_END = pd.Period('2019-12', freq='M')*

*>>>*

*>>> # Splits de data in de trainingsset*

*>>> Y\_train = filtered\_df.loc[:, :DATE\_TRAIN\_END] # Data tot en met 2019-12*

*>>>*

*>>> # Standaardiseer de trainingsdata*

*>>> Y\_train\_std = standardize(Y\_train.values.T).T*

*>>>*

*>>> # Initialiseer het model en pas PCA toe*

*>>> num\_factors = 9 # Gebruik bijvoorbeeld 9 factoren voor deze test*

*>>> model = DynamicFactorModel(Y\_train, num\_factors)*

*>>> model.std\_data = Y\_train\_std.T*

*>>> model.apply\_pca() # PCA toepassen om factoren te extraheren*

*>>>*

*>>> # Yule-Walker schatting uitvoeren*

*>>> model.yw\_estimation()*

*>>>*

*>>> # Voorspel factoren voor de volgende tijdstempel na de laatste van de trainingsset*

*>>> next\_timestamp = '2020-01' # De volgende maand na de laatste trainingsmaand*

*>>> factor\_forecast = model.factor\_forecast(next\_timestamp, scenarios=1)*

*>>>*

*>>> # Inspecteer de voorspelde factoren*

*>>> print("Predicted factors for '2020-01':\n", factor\_forecast)*

*Predicted factors for '2020-01':*

*[[-5.530608 -1.18141764 -0.33418304 -1.17280904 0.14506855 0.20184185*

*0.34628876 0.09120053 0.03231743]]*

*>>>*

## Stap 4: Voorspel variabelen op basis van factoren t+1

Nu gebruik je het getrainde ElasticNet-model en de voorspelde factoren van tijdstempel t+1om de 66 variabelen op tijdstempel t+1 te voorspellen.

Dit geeft je een nieuwe voorspelde waarde voor elke variabele in de dataset.

### CHECK

*Dit is de terminal output na check stap 4: runnen van mainPCAstatic.py:*

*PS C:\Thesis\04. Models> & C:/Users/mayac/AppData/Local/Microsoft/WindowsApps/python3.12.exe "c:/Thesis/04. Models/mainPCAstatic.py"*

*Evaluating model with 5 factors*

*Shape of factor\_forecast before transpose: (1, 5)*

*Predicted variables for 2020-01:*

*[[-0.30573665 0.25379218 0.17394232 -0.03927756 0.17468824 -0.43481686*

*-0.30174791 -0.36659139 0.3350193 -0.201414 -0.06073403 0.00964723*

*0.10923181 -0.01104068 0.36003988 0.66754804 0.12590809 0.44714363*

*0.32582807 0.46635958 0.42172514 -0.01948389 4.23809916 0.5240047*

*0.44751704 0.28602827 1.06095044 6.20614349 1.01684909 -0.19310037*

*0.81221461 0.88590477 -0.65966454 0.81175362 0.75482848 0.74017354*

*0.76261498 0.55640286 0.72223207 0.57103695 0.70198015 0.72851762*

*0.65486298 0.61108392 -1.28173351 -0.99399939 -0.8905397 -0.6535491*

*-0.06239688 -0.88254897 -0.88254897 -1.10199297 -0.2058529 -1.01858023*

*-0.92553801 -0.96992353 -1.18624003 -0.89861576 -0.53870414 0.36276693*

*-1.26884328 -0.800633 -0.90114051 -0.16067056 -0.85532542 -0.8717549 ]]*

*Evaluating model with 6 factors*

*Shape of factor\_forecast before transpose: (1, 6)*

*Predicted variables for 2020-01:*

*[[-0.30661141 0.25084453 0.1726358 -0.04111718 0.17393465 -0.43647241*

*-0.30038177 -0.37075752 0.3338386 -0.20104233 -0.06301578 0.0082263*

*0.10221985 -0.01868239 0.34966773 0.65685461 0.11830109 0.44199808*

*0.31683962 0.46350612 0.41451037 -0.02391383 4.2607088 0.51663349*

*0.44374794 0.31299064 1.09282787 6.17687941 0.99778115 -0.15981681*

*0.82343602 0.89152513 -0.65047436 0.81374399 0.75747968 0.74282155*

*0.76264347 0.55897221 0.72548365 0.57165635 0.70400065 0.72890001*

*0.65684123 0.61527671 -1.2808081 -0.99330019 -0.88995105 -0.6532884*

*-0.06185025 -0.88167567 -0.88167567 -1.10201184 -0.20568886 -1.01767952*

*-0.92371692 -0.97143785 -1.18561877 -0.89835101 -0.53875604 0.36275293*

*-1.26962147 -0.79881848 -0.90243447 -0.16036632 -0.85416908 -0.86994462]]*

*Evaluating model with 7 factors*

*Shape of factor\_forecast before transpose: (1, 7)*

*Predicted variables for 2020-01:*

*[[-3.16581788e-01 2.12793183e-01 1.52306782e-01 -5.79990102e-02*

*1.43010709e-01 -4.45006590e-01 -3.02183349e-01 -4.10515110e-01*

*3.12416282e-01 -2.07426122e-01 -1.08231850e-01 2.39094437e-03*

*-4.44167852e-03 -1.34744954e-01 1.92243491e-01 5.44053519e-01*

*7.74951313e-02 3.93608585e-01 1.72098329e-01 3.67433626e-01*

*3.08244959e-01 -1.03326300e-01 4.40597987e+00 5.13868543e-01*

*4.40343026e-01 3.07918636e-01 1.12935187e+00 6.26575493e+00*

*1.00859625e+00 -1.23367535e-01 8.41513372e-01 9.03201351e-01*

*-6.25149398e-01 8.42417440e-01 8.03548963e-01 7.88145890e-01*

*7.84965770e-01 5.85197714e-01 7.54874182e-01 5.82745111e-01*

*7.55843284e-01 7.51232113e-01 6.82347689e-01 6.79627998e-01*

*-1.24449014e+00 -9.61037287e-01 -8.51991012e-01 -6.24433484e-01*

*-4.64509101e-02 -8.43966343e-01 -8.43966343e-01 -1.05906267e+00*

*-1.86087649e-01 -9.82318621e-01 -8.96917888e-01 -9.81132056e-01*

*-1.19885089e+00 -8.86361836e-01 -5.48544928e-01 3.63863955e-01*

*-1.27624186e+00 -7.99378322e-01 -9.05095731e-01 -1.66558227e-01*

*-8.66441667e-01 -8.83059220e-01]]*

*Evaluating model with 8 factors*

*Shape of factor\_forecast before transpose: (1, 8)*

*Predicted variables for 2020-01:*

*[[-3.21104350e-01 2.13303374e-01 1.52558800e-01 -5.82785437e-02*

*1.40370743e-01 -4.45017441e-01 -3.03030167e-01 -4.09749234e-01*

*3.11334174e-01 -2.09186501e-01 -1.09153285e-01 3.57777203e-03*

*-4.73952088e-03 -1.34317114e-01 1.95057930e-01 5.36567300e-01*

*7.57427987e-02 3.88373946e-01 1.75938740e-01 3.66632686e-01*

*3.07901455e-01 -1.04939158e-01 4.41196604e+00 5.71232178e-01*

*3.47601377e-01 3.24053489e-01 1.09156058e+00 6.27100482e+00*

*1.00653638e+00 -1.10735230e-01 8.29989833e-01 9.31148484e-01*

*-5.71545223e-01 8.42279950e-01 8.02497836e-01 7.91217232e-01*

*7.83770056e-01 5.86096580e-01 7.55291638e-01 5.82498516e-01*

*7.52683656e-01 7.50265774e-01 6.82488856e-01 6.80393788e-01*

*-1.24633522e+00 -9.62610013e-01 -8.53290246e-01 -6.25930781e-01*

*-4.85066352e-02 -8.43975186e-01 -8.43975186e-01 -1.05997721e+00*

*-1.86966953e-01 -9.85147685e-01 -8.99045940e-01 -9.83638243e-01*

*-1.19831268e+00 -8.87248082e-01 -5.48276164e-01 3.64253199e-01*

*-1.28052455e+00 -7.95061734e-01 -9.05223654e-01 -1.67702710e-01*

*-8.67804379e-01 -8.87214069e-01]]*

*Evaluating model with 9 factors*

*Shape of factor\_forecast before transpose: (1, 9)*

*Predicted variables for 2020-01:*

*[[-3.21273996e-01 2.12880830e-01 1.52249377e-01 -5.88864875e-02*

*1.39898029e-01 -4.45469472e-01 -3.03476897e-01 -4.10159858e-01*

*3.11204436e-01 -2.09335727e-01 -1.09870123e-01 3.46312864e-03*

*-4.98827965e-03 -1.34630414e-01 1.94654305e-01 5.36139970e-01*

*7.68605918e-02 3.87838075e-01 1.75159033e-01 3.65781021e-01*

*3.06854666e-01 -1.05146301e-01 4.41248015e+00 5.71857422e-01*

*3.43724672e-01 3.21478274e-01 1.09156135e+00 6.27446678e+00*

*1.01500490e+00 -9.79703171e-02 8.19718137e-01 9.26612730e-01*

*-5.76038908e-01 8.42250162e-01 8.02567671e-01 7.91995167e-01*

*7.83680066e-01 5.86164115e-01 7.54980912e-01 5.82860146e-01*

*7.52902148e-01 7.50446552e-01 6.82730810e-01 6.80712912e-01*

*-1.24576327e+00 -9.62147141e-01 -8.52927373e-01 -6.25756618e-01*

*-4.83674570e-02 -8.43695390e-01 -8.43695390e-01 -1.05950700e+00*

*-1.86892141e-01 -9.84838956e-01 -8.98609353e-01 -9.83884194e-01*

*-1.19855359e+00 -8.87194103e-01 -5.49022888e-01 3.64187967e-01*

*-1.28147378e+00 -7.95507747e-01 -9.05953839e-01 -1.68309725e-01*

*-8.68814349e-01 -8.87765658e-01]]*

*Evaluating model with 10 factors*

*Shape of factor\_forecast before transpose: (1, 10)*

*Predicted variables for 2020-01:*

*[[-3.19643579e-01 2.13006548e-01 1.52440826e-01 -5.85026166e-02*

*1.39278104e-01 -4.43976173e-01 -3.04236868e-01 -4.09695127e-01*

*3.10453024e-01 -2.09412205e-01 -1.09247294e-01 1.96133810e-03*

*-5.83878326e-03 -1.34607666e-01 1.95172051e-01 5.37016444e-01*

*7.60622759e-02 3.87134760e-01 1.74141673e-01 3.65283629e-01*

*3.06439575e-01 -1.06364829e-01 4.40971334e+00 5.60883172e-01*

*3.36348873e-01 3.46798294e-01 1.08397091e+00 6.28505021e+00*

*1.05354695e+00 -7.64415057e-02 8.35820547e-01 9.37313456e-01*

*-5.76757497e-01 8.42355995e-01 8.02910219e-01 7.91896882e-01*

*7.84027322e-01 5.85764690e-01 7.54882903e-01 5.82704294e-01*

*7.53129221e-01 7.50573911e-01 6.82687538e-01 6.80662343e-01*

*-1.24603357e+00 -9.62174495e-01 -8.52814836e-01 -6.25553780e-01*

*-4.85105785e-02 -8.43959634e-01 -8.43959634e-01 -1.05957690e+00*

*-1.86785991e-01 -9.84780274e-01 -8.98633746e-01 -9.84133271e-01*

*-1.19931741e+00 -8.87952720e-01 -5.49195517e-01 3.64145452e-01*

*-1.28153304e+00 -7.97410715e-01 -9.06424594e-01 -1.68159713e-01*

*-8.69453834e-01 -8.87958495e-01]]*

*Evaluating model with 11 factors*

*Shape of factor\_forecast before transpose: (1, 11)*

*Predicted variables for 2020-01:*

*[[-3.20979249e-01 2.14343661e-01 1.53765266e-01 -5.75026393e-02*

*1.39662877e-01 -4.46103927e-01 -3.03197927e-01 -4.08360755e-01*

*3.09977448e-01 -2.09687418e-01 -1.07771555e-01 3.55055108e-04*

*-2.85501340e-03 -1.31601356e-01 2.01621912e-01 5.42343623e-01*

*7.31517851e-02 3.86715335e-01 1.81105069e-01 3.70335506e-01*

*3.12692045e-01 -1.03886427e-01 4.45380399e+00 5.87613945e-01*

*3.89668021e-01 4.49356185e-01 1.10156960e+00 6.30539940e+00*

*1.08648411e+00 -4.59626696e-02 8.75702791e-01 9.78581943e-01*

*-5.29585109e-01 8.42018274e-01 8.02029336e-01 7.90473715e-01*

*7.83251354e-01 5.84949290e-01 7.54919359e-01 5.82307970e-01*

*7.50914985e-01 7.49815014e-01 6.81580483e-01 6.77930071e-01*

*-1.25124886e+00 -9.67001885e-01 -8.58017634e-01 -6.28997134e-01*

*-5.09667237e-02 -8.48055319e-01 -8.48055319e-01 -1.06572277e+00*

*-1.89117415e-01 -9.89738091e-01 -9.01662789e-01 -9.86030074e-01*

*-1.20106800e+00 -8.90745578e-01 -5.48905132e-01 3.63611172e-01*

*-1.28431768e+00 -7.98885842e-01 -9.08917419e-01 -1.67670440e-01*

*-8.68824156e-01 -8.87057705e-01]]*

*Evaluating model with 12 factors*

*Shape of factor\_forecast before transpose: (1, 12)*

*Predicted variables for 2020-01:*

*[[-3.14013570e-01 2.15647448e-01 1.50143928e-01 -5.79792105e-02*

*1.24474731e-01 -4.40324641e-01 -3.10330885e-01 -4.05353542e-01*

*3.07786369e-01 -2.19863260e-01 -1.10312422e-01 -4.02284561e-03*

*-7.81017561e-03 -1.32475443e-01 2.09427768e-01 5.47926603e-01*

*7.10542503e-02 3.71093684e-01 1.94473076e-01 3.69020768e-01*

*3.14103049e-01 -1.13847282e-01 4.55686213e+00 6.43323989e-01*

*2.10395002e-01 3.47154895e-01 1.04602050e+00 6.38624932e+00*

*1.14072174e+00 -2.49353858e-01 7.86520083e-01 1.00673599e+00*

*-4.97538183e-01 8.41284227e-01 8.02869236e-01 7.93529543e-01*

*7.82066899e-01 5.84470398e-01 7.51283499e-01 5.80592594e-01*

*7.45603174e-01 7.48511650e-01 6.80205981e-01 6.78849559e-01*

*-1.26107002e+00 -9.74472144e-01 -8.64722172e-01 -6.35075416e-01*

*-5.76234934e-02 -8.49830205e-01 -8.49830205e-01 -1.07111298e+00*

*-1.91750704e-01 -1.00158190e+00 -9.14608791e-01 -9.90362656e-01*

*-1.20305019e+00 -8.88745673e-01 -5.46613050e-01 3.62741365e-01*

*-1.29937268e+00 -7.86863301e-01 -9.06957071e-01 -1.67600564e-01*

*-8.80161405e-01 -9.00853688e-01]]*

*Results saved to results\_PCAstatic\_with\_AIC\_BIC\_AdjustedR2\_LogLikelihood\_Residuals.xlsx*

*PS C:\Thesis\04. Models>*

## Stap 5: Maak twee nieuwe matrices

Een matrix voor de voorspelde factoren vanaf tijdstempel t+1.

Een matrix voor de voorspelde variabelen vanaf tijdstempel t+1.

### CHECK

*Dit is de terminal output na check stap 5: runnen van mainPCAstatic.py:*

*PS C:\Thesis\04. Models> & C:/Users/mayac/AppData/Local/Microsoft/WindowsApps/python3.12.exe "c:/Thesis/04. Models/mainPCAstatic.py"*

*Evaluating model with 5 factors*

*Predicted variables for 2020-01:*

*[[-0.30573665 0.25379218 0.17394232 -0.03927756 0.17468824 -0.43481686*

*-0.30174791 -0.36659139 0.3350193 -0.201414 -0.06073403 0.00964723*

*0.10923181 -0.01104068 0.36003988 0.66754804 0.12590809 0.44714363*

*0.32582807 0.46635958 0.42172514 -0.01948389 4.23809916 0.5240047*

*0.44751704 0.28602827 1.06095044 6.20614349 1.01684909 -0.19310037*

*0.81221461 0.88590477 -0.65966454 0.81175362 0.75482848 0.74017354*

*0.76261498 0.55640286 0.72223207 0.57103695 0.70198015 0.72851762*

*0.65486298 0.61108392 -1.28173351 -0.99399939 -0.8905397 -0.6535491*

*-0.06239688 -0.88254897 -0.88254897 -1.10199297 -0.2058529 -1.01858023*

*-0.92553801 -0.96992353 -1.18624003 -0.89861576 -0.53870414 0.36276693*

*-1.26884328 -0.800633 -0.90114051 -0.16067056 -0.85532542 -0.8717549 ]]*

*Evaluating model with 6 factors*

*Predicted variables for 2020-01:*

*[[-0.30661141 0.25084453 0.1726358 -0.04111718 0.17393465 -0.43647241*

*-0.30038177 -0.37075752 0.3338386 -0.20104233 -0.06301578 0.0082263*

*0.10221985 -0.01868239 0.34966773 0.65685461 0.11830109 0.44199808*

*0.31683962 0.46350612 0.41451037 -0.02391383 4.2607088 0.51663349*

*0.44374794 0.31299064 1.09282787 6.17687941 0.99778115 -0.15981681*

*0.82343602 0.89152513 -0.65047436 0.81374399 0.75747968 0.74282155*

*0.76264347 0.55897221 0.72548365 0.57165635 0.70400065 0.72890001*

*0.65684123 0.61527671 -1.2808081 -0.99330019 -0.88995105 -0.6532884*

*-0.06185025 -0.88167567 -0.88167567 -1.10201184 -0.20568886 -1.01767952*

*-0.92371692 -0.97143785 -1.18561877 -0.89835101 -0.53875604 0.36275293*

*-1.26962147 -0.79881848 -0.90243447 -0.16036632 -0.85416908 -0.86994462]]*

*Evaluating model with 7 factors*

*Predicted variables for 2020-01:*

*[[-3.16581788e-01 2.12793183e-01 1.52306782e-01 -5.79990102e-02*

*1.43010709e-01 -4.45006590e-01 -3.02183349e-01 -4.10515110e-01*

*3.12416282e-01 -2.07426122e-01 -1.08231850e-01 2.39094437e-03*

*-4.44167852e-03 -1.34744954e-01 1.92243491e-01 5.44053519e-01*

*7.74951313e-02 3.93608585e-01 1.72098329e-01 3.67433626e-01*

*3.08244959e-01 -1.03326300e-01 4.40597987e+00 5.13868543e-01*

*4.40343026e-01 3.07918636e-01 1.12935187e+00 6.26575493e+00*

*1.00859625e+00 -1.23367535e-01 8.41513372e-01 9.03201351e-01*

*-6.25149398e-01 8.42417440e-01 8.03548963e-01 7.88145890e-01*

*7.84965770e-01 5.85197714e-01 7.54874182e-01 5.82745111e-01*

*7.55843284e-01 7.51232113e-01 6.82347689e-01 6.79627998e-01*

*-1.24449014e+00 -9.61037287e-01 -8.51991012e-01 -6.24433484e-01*

*-4.64509101e-02 -8.43966343e-01 -8.43966343e-01 -1.05906267e+00*

*-1.86087649e-01 -9.82318621e-01 -8.96917888e-01 -9.81132056e-01*

*-1.19885089e+00 -8.86361836e-01 -5.48544928e-01 3.63863955e-01*

*-1.27624186e+00 -7.99378322e-01 -9.05095731e-01 -1.66558227e-01*

*-8.66441667e-01 -8.83059220e-01]]*

*Evaluating model with 8 factors*

*Predicted variables for 2020-01:*

*[[-3.21104350e-01 2.13303374e-01 1.52558800e-01 -5.82785437e-02*

*1.40370743e-01 -4.45017441e-01 -3.03030167e-01 -4.09749234e-01*

*3.11334174e-01 -2.09186501e-01 -1.09153285e-01 3.57777203e-03*

*-4.73952088e-03 -1.34317114e-01 1.95057930e-01 5.36567300e-01*

*7.57427987e-02 3.88373946e-01 1.75938740e-01 3.66632686e-01*

*3.07901455e-01 -1.04939158e-01 4.41196604e+00 5.71232178e-01*

*3.47601377e-01 3.24053489e-01 1.09156058e+00 6.27100482e+00*

*1.00653638e+00 -1.10735230e-01 8.29989833e-01 9.31148484e-01*

*-5.71545223e-01 8.42279950e-01 8.02497836e-01 7.91217232e-01*

*7.83770056e-01 5.86096580e-01 7.55291638e-01 5.82498516e-01*

*7.52683656e-01 7.50265774e-01 6.82488856e-01 6.80393788e-01*

*-1.24633522e+00 -9.62610013e-01 -8.53290246e-01 -6.25930781e-01*

*-4.85066352e-02 -8.43975186e-01 -8.43975186e-01 -1.05997721e+00*

*-1.86966953e-01 -9.85147685e-01 -8.99045940e-01 -9.83638243e-01*

*-1.19831268e+00 -8.87248082e-01 -5.48276164e-01 3.64253199e-01*

*-1.28052455e+00 -7.95061734e-01 -9.05223654e-01 -1.67702710e-01*

*-8.67804379e-01 -8.87214069e-01]]*

*Evaluating model with 9 factors*

*Predicted variables for 2020-01:*

*[[-3.21273996e-01 2.12880830e-01 1.52249377e-01 -5.88864875e-02*

*1.39898029e-01 -4.45469472e-01 -3.03476897e-01 -4.10159858e-01*

*3.11204436e-01 -2.09335727e-01 -1.09870123e-01 3.46312864e-03*

*-4.98827965e-03 -1.34630414e-01 1.94654305e-01 5.36139970e-01*

*7.68605918e-02 3.87838075e-01 1.75159033e-01 3.65781021e-01*

*3.06854666e-01 -1.05146301e-01 4.41248015e+00 5.71857422e-01*

*3.43724672e-01 3.21478274e-01 1.09156135e+00 6.27446678e+00*

*1.01500490e+00 -9.79703171e-02 8.19718137e-01 9.26612730e-01*

*-5.76038908e-01 8.42250162e-01 8.02567671e-01 7.91995167e-01*

*7.83680066e-01 5.86164115e-01 7.54980912e-01 5.82860146e-01*

*7.52902148e-01 7.50446552e-01 6.82730810e-01 6.80712912e-01*

*-1.24576327e+00 -9.62147141e-01 -8.52927373e-01 -6.25756618e-01*

*-4.83674570e-02 -8.43695390e-01 -8.43695390e-01 -1.05950700e+00*

*-1.86892141e-01 -9.84838956e-01 -8.98609353e-01 -9.83884194e-01*

*-1.19855359e+00 -8.87194103e-01 -5.49022888e-01 3.64187967e-01*

*-1.28147378e+00 -7.95507747e-01 -9.05953839e-01 -1.68309725e-01*

*-8.68814349e-01 -8.87765658e-01]]*

*Evaluating model with 10 factors*

*Predicted variables for 2020-01:*

*[[-3.19643579e-01 2.13006548e-01 1.52440826e-01 -5.85026166e-02*

*1.39278104e-01 -4.43976173e-01 -3.04236868e-01 -4.09695127e-01*

*3.10453024e-01 -2.09412205e-01 -1.09247294e-01 1.96133810e-03*

*-5.83878326e-03 -1.34607666e-01 1.95172051e-01 5.37016444e-01*

*7.60622759e-02 3.87134760e-01 1.74141673e-01 3.65283629e-01*

*3.06439575e-01 -1.06364829e-01 4.40971334e+00 5.60883172e-01*

*3.36348873e-01 3.46798294e-01 1.08397091e+00 6.28505021e+00*

*1.05354695e+00 -7.64415057e-02 8.35820547e-01 9.37313456e-01*

*-5.76757497e-01 8.42355995e-01 8.02910219e-01 7.91896882e-01*

*7.84027322e-01 5.85764690e-01 7.54882903e-01 5.82704294e-01*

*7.53129221e-01 7.50573911e-01 6.82687538e-01 6.80662343e-01*

*-1.24603357e+00 -9.62174495e-01 -8.52814836e-01 -6.25553780e-01*

*-4.85105785e-02 -8.43959634e-01 -8.43959634e-01 -1.05957690e+00*

*-1.86785991e-01 -9.84780274e-01 -8.98633746e-01 -9.84133271e-01*

*-1.19931741e+00 -8.87952720e-01 -5.49195517e-01 3.64145452e-01*

*-1.28153304e+00 -7.97410715e-01 -9.06424594e-01 -1.68159713e-01*

*-8.69453834e-01 -8.87958495e-01]]*

*Evaluating model with 11 factors*

*Predicted variables for 2020-01:*

*[[-3.20979249e-01 2.14343661e-01 1.53765266e-01 -5.75026393e-02*

*1.39662877e-01 -4.46103927e-01 -3.03197927e-01 -4.08360755e-01*

*3.09977448e-01 -2.09687418e-01 -1.07771555e-01 3.55055108e-04*

*-2.85501340e-03 -1.31601356e-01 2.01621912e-01 5.42343623e-01*

*7.31517851e-02 3.86715335e-01 1.81105069e-01 3.70335506e-01*

*3.12692045e-01 -1.03886427e-01 4.45380399e+00 5.87613945e-01*

*3.89668021e-01 4.49356185e-01 1.10156960e+00 6.30539940e+00*

*1.08648411e+00 -4.59626696e-02 8.75702791e-01 9.78581943e-01*

*-5.29585109e-01 8.42018274e-01 8.02029336e-01 7.90473715e-01*

*7.83251354e-01 5.84949290e-01 7.54919359e-01 5.82307970e-01*

*7.50914985e-01 7.49815014e-01 6.81580483e-01 6.77930071e-01*

*-1.25124886e+00 -9.67001885e-01 -8.58017634e-01 -6.28997134e-01*

*-5.09667237e-02 -8.48055319e-01 -8.48055319e-01 -1.06572277e+00*

*-1.89117415e-01 -9.89738091e-01 -9.01662789e-01 -9.86030074e-01*

*-1.20106800e+00 -8.90745578e-01 -5.48905132e-01 3.63611172e-01*

*-1.28431768e+00 -7.98885842e-01 -9.08917419e-01 -1.67670440e-01*

*-8.68824156e-01 -8.87057705e-01]]*

*Evaluating model with 12 factors*

*Predicted variables for 2020-01:*

*[[-3.14013570e-01 2.15647448e-01 1.50143928e-01 -5.79792105e-02*

*1.24474731e-01 -4.40324641e-01 -3.10330885e-01 -4.05353542e-01*

*3.07786369e-01 -2.19863260e-01 -1.10312422e-01 -4.02284561e-03*

*-7.81017561e-03 -1.32475443e-01 2.09427768e-01 5.47926603e-01*

*7.10542503e-02 3.71093684e-01 1.94473076e-01 3.69020768e-01*

*3.14103049e-01 -1.13847282e-01 4.55686213e+00 6.43323989e-01*

*2.10395002e-01 3.47154895e-01 1.04602050e+00 6.38624932e+00*

*1.14072174e+00 -2.49353858e-01 7.86520083e-01 1.00673599e+00*

*-4.97538183e-01 8.41284227e-01 8.02869236e-01 7.93529543e-01*

*7.82066899e-01 5.84470398e-01 7.51283499e-01 5.80592594e-01*

*7.45603174e-01 7.48511650e-01 6.80205981e-01 6.78849559e-01*

*-1.26107002e+00 -9.74472144e-01 -8.64722172e-01 -6.35075416e-01*

*-5.76234934e-02 -8.49830205e-01 -8.49830205e-01 -1.07111298e+00*

*-1.91750704e-01 -1.00158190e+00 -9.14608791e-01 -9.90362656e-01*

*-1.20305019e+00 -8.88745673e-01 -5.46613050e-01 3.62741365e-01*

*-1.29937268e+00 -7.86863301e-01 -9.06957071e-01 -1.67600564e-01*

*-8.80161405e-01 -9.00853688e-01]]*

*Results saved to results\_PCAstatic\_with\_AIC\_BIC\_AdjustedR2\_LogLikelihood\_Residuals.xlsx*

*Predicted factors and variables matrices saved to separate Excel files for each number of factors.*

*PS C:\Thesis\04. Models>*

## Stap 6: Hertrain het model met nieuwe voorspellingen

Gebruik nu alle data (inclusief de voorspellingen tot tijdstempel t+1) om je model opnieuw te schatten. Dit betekent dat je je factormodel (bijvoorbeeld PCA) opnieuw toepast op de geüpdatete dataset.

Voorspel de factoren en variabelen voor tijdstempel t+2.

## Stap 7: Voeg voorspellingen toe en herhaal tot tijdshorizon T

Voeg de voorspellingen voor t+2 toe aan je matrices.

Herhaal deze procedure om door te gaan met voorspellen voor toekomstige tijdstempels tot je een bepaalde tijdshorizon T hebt bereikt.

## Vraag aan chat:

* Waar in al mijn scripts voer ik stap X uit?
* Geloof je dat we stap X correct uitvoeren?
* Hoe kunnen we even kort tussendoor inspecteren of dit goed gaat? Kan ik een command in de terminal runnen?
* Klopt deze output in de terminal bij onze verwachtingen?

## Uitleg aan chat:

hoi chat! Ik ben al best ver met het schrijven van mijn master scriptie econometrie code in python, maar ik heb je hulp nodig bij een nieuw probleem. Zal ik je mijn probleem uitleggen, en mijn huidige scripts met je delen?

Alright! Ik onderzoek of de voorspellende kracht van een dynamic factor model verbeterd als je Partial Least Squares gebruikt om de factoren te extraheren uit de data, versus als je Principal Component Analysis gebruikt. Verder onderzoek ik of de voorspellende kracht verbetert als je niet alleen statische variabelen gebruikt als input data (GDP, CPI, etc.) versus als je een aantal forward looking indicators hieraan toevoegd. Even ter info: ik heb 66 statische variabelen die ik wil voorspellen, de 29 forward looking variables worden hieraan toegevoegd om de 'combined dataset' te vormen, maar deze forward looking variables wil ik verder niet voorspellen, dus gebruik deze alleen als input voor de factoren. Dan heb ik in totaal maandelijke data van 1995-01 tot en mt 2023-11, deze data heb ik opgesplits in train data van 1995-01 tot en met 2019-12, en een validatieset vanaf 2020-01 tm 2023-11. Deze validatie dataset wil ik echt even apart houden om later out of sample mee te testen. De train data set wil ik graag met de 80%-20% regel opsplitsen, zodat we het getrainde model ook in sample kunnen testen. Snap je uit mijn verhaal dat ik 4 modellen met elkaar vergelijk? PCAstatic, PCAcombined, PLSstatic en PLScombined?